Anamoly Detection in SAR Imagery

Love Sachdeva   
Computer Science And Engineering  
Vellore Institute of TechnologyChennai , India  
Love.sachdeva2019@vitstudent.ac.in

Abhishek Yadav   
Computer Science And Engineering  
Vellore Institute of TechnologyChennai , India  
Abhishek.yadav2019@vitstudent.ac.in

Deb Shankar Modal   
Computer Science And Engineering  
Vellore Institute of TechnologyChennai , India  
Debshankar.mondal2019@vitstudent.ac.in

*Abstract*—

Anamoly detection in high-goal synthetic gap radar (SAR) imagery is a principal and provoking issue because of the perplexing conditions. In this paper, a RetinaNet-Plus technique is introduced for transport detection in high-goal SAR imagery dependent on RetinaNet network altered. Furthermore, central misfortune is utilized to address the class unevenness and to expand the significance of the hard models during preparing. The analyses on SAR transport SSDD dataset and TerraSAR-X image from Barcelona port, show that our strategy is more exact than the current calculations and is successful for transport detection of high-goal SAR imagery.

INDEX TERMS : Anamoly , SAR Imagery , RetinaNet , dataset

# **Introduction**

Synthetic Aperture Radar (SAR) disaster monitoring , all-climate abilities and it enjoys the benefits of entering cloud downpour and autonomous of the light source contrasted and conventional optical imaging innovation. As of late, with the fast improvement of imaging innovation in the space of remote detecting, many satellites and ethereal sensors furnish SAR imagery with high-goal. These images work with a wide scope of uses in the field of safeguard, ecological administration, country asset investigation, and catastrophic event observing, where targets are regularly involved [3]. As a major problem looked for remote detecting image examination, object detection in remote detecting images assumes a significant part for both military and regular citizen applications. Notwithstanding, it's as yet a moving problem because of the scale variety, visual particularity, little objective problem, multi-directional problem and foundation intricacy of high-goal remote detecting images. Moreover, images in amount and quality make an incredibly high computational expenses, which additionally expands the troubles of item detection for close constant applications. As of late, Ship detection is a significant point in the field of SAR. Ship detection is an intricate problem, requiring the arrangement of two primary assignments. In the first place, the indicator should take care of the acknowledgment problem, to recognize forefront ships from the foundation and dole out them the appropriate article class names. Second, the locator should take care of the localization problem, to allot precise bouncing boxes to various ships.

Lately, with the quick improvement of profound learning, analysts in the PC vision field have made significant leap forwards in object detection. As of now, the standard item detection algorithms are basically founded on profound learning models, which can be partitioned into two classes:two-stage detection algorithm, what isolates the detection problem into two phases, first producing locale proposition that channels the vast majority of the negative examples, then, at that point, competitor area order (by and large should be refined for area). Regular instances of such algorithms are RCNN algorithms dependent on locale proposition, for example, R-CNN , Fast R-CNN , Faster R-CNN, Mask-RCNN,etc.; one-stage detection algorithm, which doesn't need the district proposition stage, straightforwardly creates the class likelihood and position coordinate upsides of the article, and analyzes ordinary algorithms, for example, YOLO [10] and SSD. Two-stage has the best article detection results, however onestage is quicker. Both of these techniques have the problem of class lopsidedness. The two-stage settles the class lopsidedness through Selective Search , EdgeBoxes, DeepMask and RPN. Generally, the quantity of competitor districts is 1-2k, and a proper forefront to-foundation proportion or online hard model mining are performed to keep up with the harmony among frontal area and foundation in the characterization. The one-stage technique has 100k distinctive size areas. Because of most of the foundation that is effortlessly recognized during preparing, this can cause preparing failure. Along these lines, the reference proposes Focal misfortune and applies it to the one-stage strategy to take care of the outrageous class lopsidedness problem through the misfortune work. Objects in huge scope SAR imagery are somewhat little in size and show up in thickly appropriated gatherings. For ships, the states of a similar ship are multi-scale because of the impact of different goals, and ships with different shapes show diversely in a similar goal SAR imagery. Thusly, it is important to think about the difference of ship scales. To construct undeniable level semantic element maps at all scales, we apply a component pyramid structure (FPN) as a spine network for highlight extraction. FPN utilizes a hierarchical design to meld the component of various goals from a solitary scale input, which further develops exactness with minimal expense.

In this paper, we present a RetinaNet-Plus for ship detection in high-goal SAR imagery. RetinaNet-Plus has three parts: a spine network for include extraction and two sub-organizations (one for arrangement and the other for box relapse). In SAR ship detection dataset (SSDD), there are some thick ship objects. Delicate NMS is brought into our structure for staying away from loss of accuracy because of set the score for adjoining area recommendations to zero as in NMS, which further develops the detection execution of thick ships. Besides, central misfortune is utilized to address the class awkwardness and to build the significance of the hard models during preparing. At last, we contrast and state-ofthe-workmanship profound CNN based strategies, show the adequacy of the proposed strategy on SSDD dataset. To more readily assess our proposed technique, one TerraSAR-X imagery from Barcelona port are utilized to test the heartiness.

# **RELATED WORK**

Present state of-the-art object detectors depend on a two-stage, proposition driven instrument. As advocated in the R-CNN system [11], the primary stage produces a scanty arrangement of competitor object areas and the subsequent stage characterizes every applicant area as one of the forefront classes or as foundation utilizing a convolutional neural organization. Through an arrangement of advances [10, 27, 19, 13], this two-stage structure reliably accomplishes top exactness on the difficult COCO benchmark [20]. In spite of the accomplishment of two-stage indicators, a characteristic inquiry to pose is: could a basic one-organize finder accomplish comparative exactness? One phase finders are applied over a normal, thick inspecting of item areas, scales, and perspective proportions. Late work on one-stage finders, for example, YOLO [25, 26] and SSD [21, 9], shows promising outcomes, yielding quicker identifiers with exactness inside 10-40% comparative with cutting edge two-stage strategies.

This paper pushes the wrap further: we present an onestage article identifier that, interestingly, matches the 2017 IEEE International Conference on Computer Vision 2380-7504/17 $31.00 © 2017 IEEE DOI 10.1109/ICCV.2017.3242999 best in class COCO AP of more intricate two-stage finders, for example, the Feature Pyramid Network (FPN) [19] or Mask R-CNN [13] variations of Faster R-CNN [27]. To accomplish this outcome, we distinguish class lopsidedness during preparing as the primary hindrance obstructing one-stage finder from accomplishing cutting edge exactness and propose another misfortune work that takes out this boundary. Class unevenness is tended to in R-CNN-like indicators by a two-stage course and inspecting heuristics. The proposition stage (e.g., Selective Search [34], EdgeBoxes [37], DeepMask [23, 24], RPN [27]) quickly limits the quantity of up-and-comer object areas to a modest number (e.g., 1-2k), sifting through most foundation tests. In the subsequent grouping stage, examining heuristics, for example, a proper forefront to-foundation proportion (1:3), or online hard model mining (OHEM) [30], are performed to keep a reasonable harmony among frontal area and foundation.

Conversely, a one-stage finder should deal with a much bigger arrangement of up-and-comer object areas consistently inspected across an image. Practically speaking this regularly sums to listing ∼100k areas that thickly cover spatial positions, scales, and perspective proportions. While comparable inspecting heuristics may likewise be applied, they are wasteful as the preparation method is as yet overwhelmed by handily ordered foundation models. This shortcoming is an exemplary problem in object detection that is regularly tended to by means of strategies, for example, bootstrapping [32, 28] or hard model mining [36, 8, 30]. In this paper, we propose another misfortune work that goes about as a more successful option in contrast to past approaches for managing class lopsidedness. The misfortune work is a powerfully scaled cross entropy misfortune, where the scaling factor rots to zero as trust in the right class builds, see Figure 1. Naturally, this scaling variable can consequently down-weight the commitment of simple models during preparing and quickly center the model around hard models. Tests show that our proposed Focal Loss empowers us to prepare a high-precision, one-stage indicator that essentially beats the choices of preparing with the examining heuristics or hard model mining, the past state-ofthe-workmanship procedures for preparing one-stage finders. At long last, we note that the specific type of the central misfortune isn't vital, and we show different launches can accomplish comparative outcomes.

To show the adequacy of the proposed central misfortune, we plan a straightforward one-stage object indicator called RetinaNet, named for its thick testing of article areas in an info image. Its configuration includes a proficient in-network highlight pyramid and utilization of anchor boxes. It draws on an assortment of late thoughts from [21, 6, 27, 19]. RetinaNet is proficient and exact; our best model, in light of a ResNet-101-FPN spine, accomplishes a COCO test-dev AP of 39.1 while running at 5 fps

.

# **PROPOSED SOLUTION**

We’ll be using Retinanet by facebook for our anomaly detection.

RetinaNet is a solitary, bound together organization made out of a

spine organization and two assignment explicit subnetworks. The

spine is liable for figuring a convolutional feature map over a whole information picture and is an off oneself

convolutional network. The first subnet performs convolutional object order on the spine's result; the

second subnet performs convolutional bouncing box regression. The two subnetworks include a basic plan that we

propose explicitly for one-stage, thick location, see Figure 3. While there are numerous potential decisions for the subtleties

of these parts, most plan boundaries are not particularly touchy to correct qualities as displayed in the trials.

We depict every part of RetinaNet next.

## Feature Pyramid Network Backbone:

We embrace the Feature Pyramid Network (FPN) from as the spine

network for RetinaNet. In short, FPN expands a standard convolutional network with a hierarchical pathway and

horizontal associations so the organization proficiently develops a

rich, multi-scale highlight pyramid from a solitary goal

input picture. Each level of the pyramid

can be utilized for distinguishing objects at an alternate scale. FPN

improves multi-scale expectations from completely convolutional

networks (FCN), as shown by its benefits for RPN also DeepMask-style proposition, too at two-stage

finders, for example, Fast R-CNN or Mask R-CNN.

Following, we construct FPN on top of the ResNet architecture. We build a pyramid with levels P3

through P7, where l shows pyramid level (Pl has resolution 2l lower than the info). As in all pyramid levels

have C = 256 channels. Subtleties of the pyramid for the most part

follow with a couple humble differences.2 While numerous

plan decisions are not vital, we stress the utilization of the

FPN spine is; starter tests utilizing highlights

from just the last ResNet layer yielded low AP.

## Focal Loss

RetinaNet has a place with the one-stage object locators in

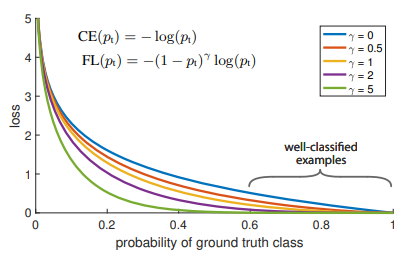
profound learning. Furthermore, During the preparation, the class lopsidedness and inconsistent commitment of hard and simple guide to the misfortune affect the detection

exactness in the one-stage object detection situation. To

counter this, the focal misfortune is proposed in reference [16]. It puts more accentuation on hard model and spotlights on the way that the deficiency of hard model is higher during preparing. It is communicated as been utilized to further develop the detection exactness.

The focal loss function is defined as

FL(pt) = −(1 − pt) γ log(pt)[17]



Focal loss image taken from original keras retinanet paper[17]

# **Experimental setup**

## Dataset

The data set that we’ve used for our model is of SAR imagery of ships. Anomaly for SAR imagery can be many things like an unidentified ship in international waters or an unknown aircraft in the airport. But to keep the dataset short due to lack of physical power in our machines so we’ll be detecting ships in the waters and identifying them as anomaly.

Title: Ship Detection from Aerial Images

Uploaded by: Larxel

Version: 1

License: CC0: Public Domain

https://www.kaggle.com/andrewmvd/ship-detection/version/1

Photo by Shaah Shahidh on Unsplash.

We also took some more images from random raw datasets containing ship SAR imagery and converted them to pascal voc type to train our model but the major dataset was from the one mentioned above.

## Software and hardware required

Software that we’ve used for this work is jupyter notebook on conda.

For hardware, we recommend a high power gpu for faster results and for the model to run all it’s epochs.

To use this hardware with the notebook, we need tensorflow, keras, and Nvdia CUDA Toolkit.

## Deep learning model

Since Fizyr had already implemented retinanet in an excellent way and made it public for free use don’t need to implement retinanet from scratch. We can just clone their github repository in our machine.

We can make a few necessary changes onto this proposed model for it to use our dataset and fine tune it to respond to our require query images. Also retinanet is more famous for usage of COOC dataset but Fizyr’s retinanet implementation allows user to use from a variety of dataset type that they are comfortable with. All though all images in the dataset must be of one dataset type only.

For our model we’ll be using Pascal Voc dataset type with annotations of the correct anchor boxes of the image in xml extensions.

https://github.com/fizyr/keras-retinanet

# **IMPLEMENTATION RESULTS**

Few results obtained using the proposed model:

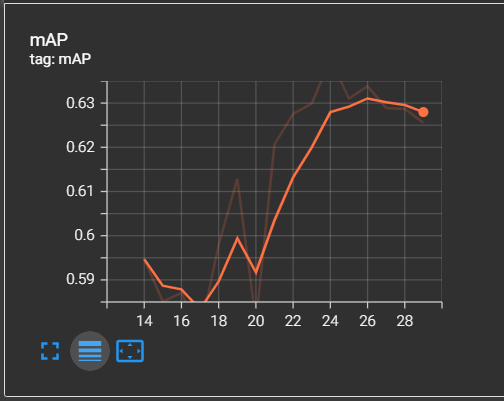




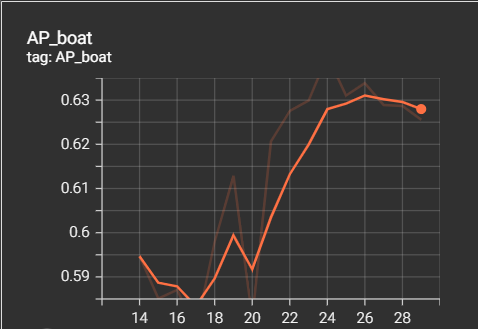




The mean Average Precision of the model with respect to the epochs is shown below. For our model we’ve stopped the training at 26th epoch and converted it into inference model for testing to avoid overfitting of the data.



Since anomalies can of be any type, we can get the mAP values for all the object types separately too. For this model since we have only taken anomalies in water bodies, the AP of the boat/ship is the same as model’s mAP.



These mAP graphs were drawn at the training time with the use of tensorboard.

# **conclusion and future works**

In this paper, we proposed a RetinaNet based strategy for anomaly detection in high-goal SAR images. Besides, central misfortune break the class irregularity and inconsistent commitment of hard and simple guides to the misfortune. The outcomes show that our system is more vigorous to high resolution SAR images than YOLOv2 and other one stage detectors. In the interim, our technique accomplished better execution for identifying ships with huge scope and high-goal SAR imagery. We can add data of aircrafts, vehicles, land faults and feed it to the model for it to be more vigorous toward all types of anomalies. Accordingly, it has extraordinary potential for wide field application. To work on the precision and decrease the work to plan the anchor boxes, we will propose an overall way to deal with advance anchor encloses for ship detection what's to come

# **REFERENCES**

1. Chang Y L, Anagaw A, Chang L, et al. Ship Detection Based on YOLOv2 for SAR Imagery[J]. Remote Sensing, 2019, 11(7): 786.
2. Deng Z, Sun H, Zhou S, et al. Multi-scale object detection in remote sensing imagery with convolutional neural networks[J]. ISPRS
3. Journal of Photogrammetry and Remote Sensing, 2018, 145: 3-22.
4. Shahzad M, Maurer M, Fraundorfer F, et al. Buildings Detection in VHR SAR Images Using Fully Convolution Neural Networks[J]. IEEE transactions on geoscience and remote sensing, 2019, 57(2): 1100-1116.
5. Ren, Shaoqing, et al. "Faster R-CNN: towards real-time object detection with region proposal networks." International Conference on Neural Information Processing Systems , pp. 91-99, 2015.
6. He, Kaiming, et al. "Mask R-CNN." IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 99, pp. 1-1, 2017.
7. Lin, Tsung-Yi, et al. "Feature Pyramid Networks for Object Detection." CVPR. Vol. 1. No. 2. 2017.
8. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
9. Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.
10. Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE
11. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.
12. Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, pp. 21-37, 2016.
13. Bodla N, Singh B, Chellappa R, et al. Soft-NMS--Improving Object Detection With One Line of Code. Proceedings of the IEEE International Conference on Computer Vision. 2017: 5561-5569.
14. Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision, pp. 1440-1448, 2015.
15. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll´ar, and C Lawrence Zitnick, “Microsoft coco: Common objects in context,” in European conference on computer vision. pp. 740–755, 2014.
16. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14; pp. 1–6, 2017.
17. Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
18. J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W.Smeulders. Selective search for object recognition. IJCV, 2013. 2, 4.
19. C. L. Zitnick and P. Doll´ar. Edge boxes: Locating object proposals from edges. In ECCV, 2014. 2.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published**